# Metadata for building the MultiMedia Patch Quilt

Vipul Kashyap[1,3] Kshitij Shah[2,3] Amit Sheth[1]
[1]LSDIS, Department of Computer Science, University of Georgia, 415 GSRC, GA 30602-7404 (http://www.cs.uga.edu/LSDIS/)
[2]Bellcore, 444 Hoes Lane, Piscataway, NJ 08854
[3]Department of Computer Science, Rutgers University, New Brunswick, NJ 08903

**Abstract.** Huge amounts of data available in a variety of digital forms has been collected and stored in thousands of repositories. However, the information relevant to a user or application need may be stored in multiple forms in different repositories. Answering a user query may require correlation of information at a *semantic level* across multiple forms and representations. We present a three-level architecture comprising of the *ontology, metadata* and *data* levels for enabling this correlation. Components of this architecture are explained by using an example from a GIS application.
Metadata is the most critical level in our architecture. Various types of metadata developed by researchers for different media are reviewed and classified *wrt* the extent they model data or information content. The reference terms and the ontology of the metadata are classified *wrt* their dependence on the application domain. We identify the type of metadata suitable for enabling correlation at a semantic level. Issues of metadata extraction, storage and association with data are also discussed.

## 1 Introduction

In recent years, huge amounts of digital data in a variety of structured, unstructured (e.g., image) and sequential (e.g., audio) formats has been collected and stored in thousands of repositories. Significant advances in managing textual, image, audio, and video databases support efficient storage and access of data of a single type in a single repository. Affordable multimedia systems and a variety of internetworking tools (including the current favorite, the World Wide Web [BL+92]) allow creation of multimedia documents, and support locating, accessing and presenting such data.

However, information relevant to a user or application need may be stored in multiple forms (e.g., structured data, image, audio, video and text) in different repositories. Answering a user query typically requires correlation of such information across multiple forms and multiple representations. Correlating various pieces of information at the physical level by pre-analysis and establishing explicit hypertext/hyper-media links is not an attractive option. For example, there could be thousands of objects that could be recognized in an image, but linking all of these objects to relevant textual or structured data would be a very time consuming and unrewarding process.

We believe that it is necessary to represent semantic information to support the correlation of heterogeneous types of information. Humans are able to abstract information efficiently from images, video or audio data displayed on the computer. This enables them to correlate information at a higher *semantic level* with other forms of representation such as the symbolic representation of data in structured databases. This capability of correlating information at a semantic level across different representations such as symbolic, image, audio and video is lacking in current multimedia systems, and has been characterized as a "semantic bottleneck" [Jai94] – a problem that we are working on. Among recent examples of visual information management in a semantically meaningful manner is the VIMSYS approach to support semantic queries on images [GWJ91]. Chu et al. [CLT94] have also adopted a semantic modeling approach for content-based retrieval for medical images.

In the InfoQuilt project, we visualize the related information in heterogeneous media types as a "patch quilt" of digital data. To enable correlation of information across heterogeneous digital media types at a semantic level, we propose a three-level architecture (Figure 1). The three main levels of this architecture are described below.

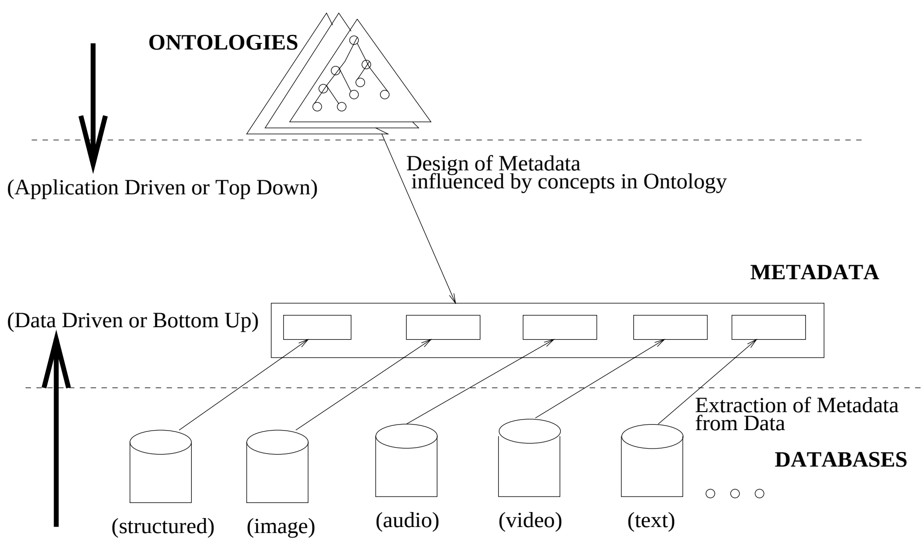


**Fig. 1.** Three level architecture for information correlation in Digital Media

**Ontologies:** These refer to the terminology/vocabulary which characterizes the content of information in a database irrespective of the media type. We shall capture this vocabulary in a symbolic representation (as opposed to, e.g., an image or audio representation). The vocabulary in general shall contain both domain-independent and domain-specific terms. The domain-independent terms may or may not depend on the characteristics of the media type.

**Metadata:** These represent *information* about the data in the individual databases and can be seen as an extension of the concept of schema in structured databases. They may describe, or be a *summary* of the information content of the individual databases in an intensional manner. They typically represent constraints between the individual media objects which are implicit and not represented in the databases themselves. Some metadata may also capture content-independent information like *location* and *time of creation.* Typically, however, metadata would capture content-dependent information like *the relief of a geographical area.*

**Data:** This is the actual (raw) data which might be represented in any of the media types. Examples of what we consider media types are structured data (data in relational or object-oriented databases), textual data, images (maybe of different modalities like X-Ray, MRI scan), audio (maybe of different modalities like monoaural, stereophonic) and video.

From the perspective of answering queries which require correlation of heterogeneous data in different repositories, the most critical level of the above architecture is the **metadata level**. For enabling semantic correlation, the metadata should be able to model the *semantics* of the data. Semantics of an object include both the "meaning" and "use" of an object [KS94b]. Researchers in the area of multidatabases have investigated the issues of semantic heterogeneity [She91] and the issues of semantic similarity and structural differences [SK92]. To compare and combine information from the various media types, we need to view them independent of their representation medium [Jai94]. The metadata level represents the level at which we shall view the information from the various media types and compare and combine them.

Some of the significant recent work in developing metadata for digital media is compiled in [KS94c]. Böhm and Rakow have provided a classification of metadata in the context of multimedia documents [BR]. Jain and Hampapur have characterized video metadata and its usage for content-based processing [JH]. Kiyoki et al. have used metadata to provide associative search of images for a set of user-given keywords [KKH]. Anderson and Stonebraker have developed a metadata schema for management of satellite images [AS]. Grosky et al. have discussed a data model for modeling and indexing metadata and providing the definition of higher abstractions [GFS]. Glavitsch et al. have demonstrated how a common vocabulary suffices to develop metadata for integration of speech and text documents [GSW]. Chen et al. describe automatic generation of metadata to support mixed media access [CHK+].

In this chapter, we have reviewed the work done by the researchers on different media types and classified the metadata designed by them. The criteria we use to classify the metadata is the extent to which they are successful in capturing the data and information content of the documents represented in various media types. The level of abstraction at which the content of the documents is captured is very important. As suggested by others (e.g. Wiederhold [Wie94] and Gruber [Gru93]), we believe that to capture the content at a level of abstraction closer to that of human beings, it is important for the metadata to model *application domain-specific* information.

It is in this context that the terms/vocabulary used to design the metadata assumes special significance. We believe that in order for the metadata to model information at a level of abstraction closer to human beings the choice of terms to construct the metadata should be domain-specific. The terms chosen should be influenced by the application in mind or user needs. Information of this type is captured at the ontology level. We categorize the vocabulary based on whether the terms are data or application-driven and whether they are domain-dependent or domain-independent.

An important component in being able to query across multiple, heterogeneous representations of related information is to be able to design and store associations of the metadata with the actual data stored in the databases. This might mean relating domain-specific terms in the metadata (e.g., cloud-cover in an image) to media-specific domain-independent terms characterizing the data (e.g., color, texture, shape of image objects). Issues of metadata extraction and storage are also discussed.

We have presented a three-level architecture to support correlation of information stored in different digital media at a higher semantic level. We review the state of art in different digital media and analyze the types of metadata used and the vocabulary from which they are constructed. We identify the types of metadata and the nature of vocabulary required to achieve semantic correlation. Where appropriate, we shall discuss examples from a GIS application to illustrate components of our three level architecture.

The organization of this chapter is as follows. We discuss issues related to the vocabulary/ontology from which the metadata are constructed in Section 2. Issues related to construction, design, storage and extraction of metadata are discussed in Section 3. Issues related to the association of data with metadata are discussed in Section 4. Conclusions and future research directions are presented in Section 5.

## 2 Characterization of the Ontology

We believe that for effective computer-based *correlation of related information* between heterogeneous digital media, it will be necessary to take advantage of knowledge pertaining to the application domain. The key to utilizing the knowledge of an application domain is identifying the basic *vocabulary* consisting of terms (or concepts) of interest to a typical user in the application domain and the interrelationships among the concepts in the ontology.

In the course of collecting a vocabulary or constructing an ontology for information represented in a particular media type, some concepts or terms may be independent of the application domain. Some of them may be media-specific while others might be media-independent. There might be some application-specific concepts for which interrelationships may be represented. They are typically independent of the media of representation.

Information represented using different *media types* can be associated with application-specific concepts and then be appropriately correlated. This forms the basis for a *semantic correlation* of information stored in heterogeneous repositories.

### 2.1 Terminological Commitments: Constructing an Ontology

An ontology may be defined as the specification of a representational vocabulary for a shared domain of discourse which may include definitions of classes, relations, functions and other objects [Gru93]. We assume that media types presenting *related information* share the same domain of discourse. However, different database designers might use different terminology for the identification and representation of various concepts. There have to be agreements on the terms used by the different designers. These agreements can be the basis of the construction of the ontology and are called *ontological commitments*. We view ontological commitments as a very important requirement for domain-dependent terms.

Typically there may be other terms in the vocabulary which may not be dependent on the domain and may be media-specific. Further it may be necessary to translate between descriptive vocabularies that involve approximating, abstracting or eliminating terms as a part of the negotiated agreement reached by the various designers. It may also be important to translate domain-dependent terms to domain-independent media-specific terms by using techniques specialized to that media type.

Research in the area of Multidatabases has typically used concept hierarchies to represent the vocabulary of the domain of discourse [YSDK91, TC94, MS95, MS94]. Let us consider the domain of a GIS application. An important problem in the GIS area is Site Location and Planning. We illustrate fragments of the concept hierarchies used in developing the ontology in Figure 2.
In the process of construction, we view the ontology from the following two different perspectives.

- The *data-driven vs* the *application-driven* perspectives.
  **Data-driven approach:** This refers to the concepts and relationships designed by interactive identification of objects in the related information stored in the databases corresponding to different media types.
  **Application-driven approach:** This refers to the concepts and relationships inspired by the class of queries for which the related information in the various media types is processed. The concept Rural Area in Figure 2 is an example of a concept obtained from the application-driven approach.
- The *domain-dependent* and the *domain-independent* perspectives.
  **Domain-dependent perspective:** This represents the concepts which are closely tied to the domain of the application we wish to model. Most of the concepts are likely to be identified using the application-driven approach.

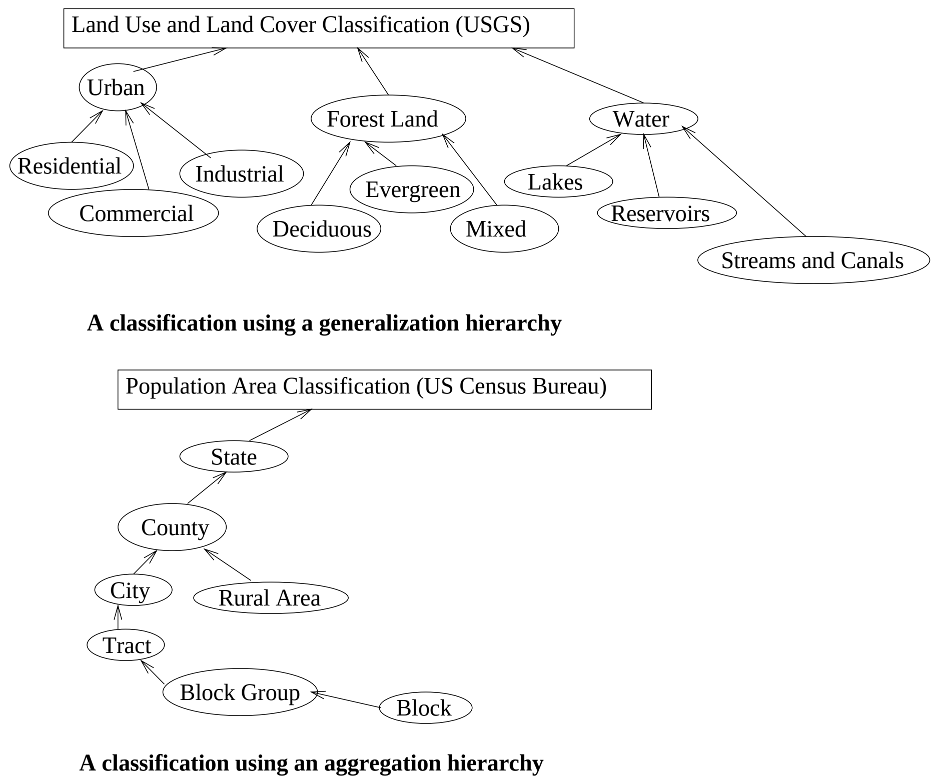


**Fig. 2.** Examples of Generalization and Aggregation hierarchies that may be used for Ontology construction

**Domain-independent perspective:** This represents the concepts required by the various media types (e.g., color, shape and texture for images, such as R-features [JH]) to identify the domain-specific concepts. These are typically independent of the application domain and are generated by using the data-driven approach.

### 2.2 Controlled Vocabulary for Digital Media

In this section we review the discussion in [KS94c] on extracting metadata. We focus on the terminology and vocabulary identified by the various researchers for characterizing the information content of the data represented in a particular media type. We identify how the various terms relate to the perspectives discussed above.

Jain and Hampapur [JH] have used domain models to assign a qualitative label to a feature (such as *pass*, *dribble* and *dunk* in basketball) and are called Q-Features. Features which rely on low level domain-independent models like object trajectories are called R-Features. We consider Q-Features as an example

| Vocabulary Feature | Media Type | Domain Dependent or Independent | Appl ication or Data Driven |
|---|---|---|---|
| Q-Features (Jain and Hampapur) | Video, Image | Domain Dependent | Application Driven |
| R-Features (Jain and Hampapur) | Video, Image | Domain Independent | Data Driven |
| English Words (Kiyoki et al.) | Image | Domain Dependent | Data Driven |
| ISCC and NBS colors (Kiyoki et al.) | Image | Domain Independent | Data Driven |
| AVHRR features (Anderson and Stonebraker) | Image | Domain Independent | Data Driven |
| NDVI (Anderson and Stonebraker) | Image | Domain Dependent | Data Driven |
| Subword units (Glavitsch et al.) | Audio, Text | Domain Dependent | Data Driven |
| Keywords (Chen et al.) | Image, Audio Text | Domain Dependent | Application and Data Driven |

**Table 1.** Controlled Vocabulary for Digital Media

of the domain-dependent, application-driven perspective and R-Features as an example of the domain-independent, data-driven perspective.

Kiyoki et al. [KKH] have used 850 basic words from the "General Basic English Dictionary" as features which are then associated with the images. We consider these features as examples of domain-dependent, data-driven perspective. They also use the color names defined by ISCC (Inter-Society Color Council) and NBS (National Bureau of Standard) as the features. We consider these as examples of the domain-independent, data-driven perspective.

Anderson and Stonebraker [AS] model some features that are primarily based on the measurements of the five channels of the Advanced Very High Resolution Radiometer (AVHRR) sensor. Other features refer to spatial (latitude, longitude) and temporal (begin date, end date) information. We consider these examples of domain-independent, data-driven perspective. However there are features like the normalized difference vegetation index (NDVI) which are derived from different channels. We consider this as an example of the domain-dependent, data-driven perspective.

Glavtisch et al. [GSW] have determined from experiments that good indexing features lay between phonemes and words. They have selected three special types of subword units: VCV-, CV- and VC-. The letter V stands for a maximum sequence of vowels and C for a maximum sequence of consonants. They process a set of speech and text documents to determine a vocabulary for the domain. The same vocabulary is used for both the speech and text media types. We consider these as examples of the domain-dependent, data-driven perspective.

Chen et al. [CHK+] use the keywords identified in the text and speech documents as their vocabulary. They have discussed issues of restricted *vs* unrestricted vocabulary. If the set of keywords is fixed, metadata based on keyword locations can be pre-computed and stored later for indexing. In general, unrestricted vocabulary produces better results than restricted vocabulary searching. They support the search by spotting keywords "on the fly". We consider these as examples of the domain-dependent, data and application-driven perspectives.

A summary of the above discussion is presented in Table 1.

The summary of the vocabulary used in digital media illustrated in Table 1 does not bring out the contribution of the media type in determining the features to characterize the information content of the databases. We recognize the fact that constructing a vocabulary with domain-independent perspective would entail both *media-specific* and *media-independent* features. We may also be able to write programs for automatic extraction of some media-specific features from the digital data. We discuss metadata extractors in further detail in Section 3.3. The role played by the media types is illustrated in Figure 3.

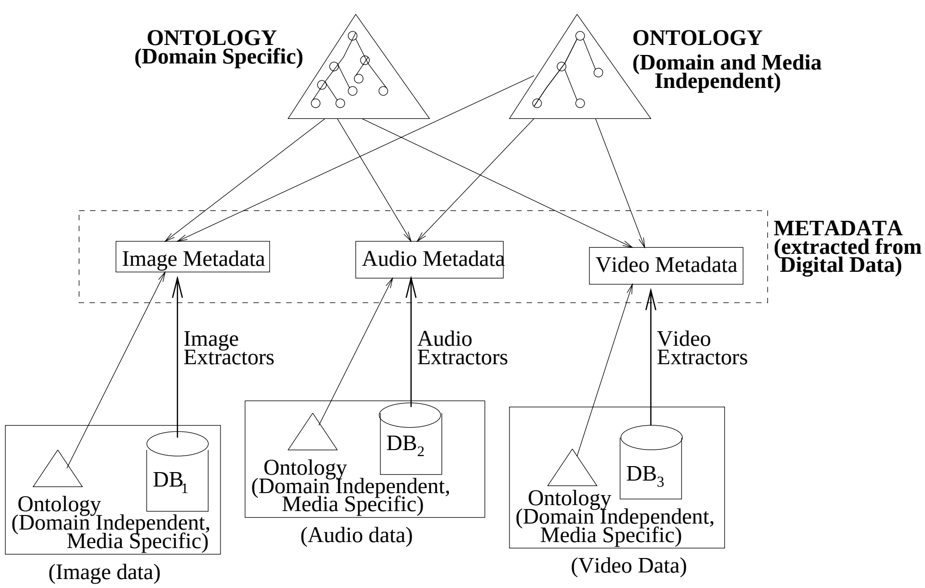


**Fig. 3.** Role of the Media Type in determining the metadata features

### 2.3 Better understanding of the query

When the terms used in a user's query are not expressive enough, or cannot be mapped by the system to the ontological concepts, the user may guide the construction of the query metadata with the help of the ontology (that may be graphically displayed). The query metadata may typically represent application-specific constraints which the answer should satisfy. We assume the representation of metadata as a collection of meta-attributes and values. For the discussion in this chapter see [KS94b, KS95] for further details.

Let us consider the Site Location and Planning Problem referred to earlier. This requires correlation of related information represented in two media types, structured databases and images. A typical query that may be asked by a decision maker trying to determine a desirable location of a shopping mall is:

Get all blocks with a population greater than five hundred with an average income greater than 30,000 per annum, that have moderate relief with a large contiguous rectangular area and are of an urban type of land use.

The metadata for the query can be constructed as follows. Let the variable X refer to the final output unified with the regions in which a mall may be built in a geographical region characterized above.

```
[ (region X) (population [> 500]) (contiguous-area [large])
(relief [moderate]) (average-income [> 30,000])
(shape [rectangular]) (land-use [Urban]) ]
```

These metadata are designed from the domain-specific ontology as its basis, and are later described as *content-descriptive domain-dependent metadata*. The current state of art in multimedia databases does not support querying at this level of abstraction. In Section 3 we survey the state of art in this area and propose the research efforts required to support the above level of abstraction.

### 2.4 Ontology guided Extraction of Metadata

The extraction of metadata from the information in various media types can be primarily guided by the domain-specific ontology though it may involve terms in the domain-independent ontology. Both content-dependent and content-independent metadata may be extracted (as discussed in Section 3).

Let us consider again the GIS application discussed earlier in the section. For this application we model the metadata as a collection of attribute value pairs. The meta-attributes can be derived from concepts in the domain-specific ontology. The values of the meta-attributes can be the set of constraints imposed on the class of domain objects represented by the set of regions. These constraints are qualitative descriptions of the spectral (i.e., color, intensity), morphological (i.e., shape related) and textural properties of the objects. For example, the meta-attribute *vegetation* can have as values, a set of image regions with the qualitative spectral constraint *green* and with the qualitative textural constraint of being either *grass-like*, *forest-like* or *shrub-like*.

Kiyoki et al. [KKH] describe the automatic extraction of *impression vectors* based on English words or ISCC and NBS colors. The users when querying an image database then use English words to query the system. One way of guiding the users could be to display the list of English words used to construct the metadata in the first place. However, because this is inconvenient the vocabulary is typically not displayed to the user.

Glavitsch et al. [GSW] describe the construction of a speech feature index for both text and audio documents based on a common vocabulary consisting of subword units as discussed earlier. Given a query, the features can be evaluated easily because the canonical forms of the features (vowel and consonant sequences like VCV-, CV- VC-) are well defined.

Chen et al. [CHK$^+$] describe the construction of keyword indices, topic change indices and layout indices. These typically depend on the content of the documents and the vocabulary is dependent on the keywords present in the documents. A query can be a set of spoken keywords which might result in the retrieval of documents containing those keywords.

In the above cases, the vocabulary is not pre-defined and depends on the content of the documents in the collection. Also, the interrelationships between the terms in the ontology is not identified. We believe that identifying these relationships would result in the reduction of the size of the vocabulary. A typical set of relationships have been identified in [MS94]. A controlled vocabulary with terms and their interrelationships can be exploited to create metadata which model domain-dependent relationships as illustrated in the case of the GIS application discussed earlier in this section.

## 3 Construction and Design of Metadata

In this section, we identify and classify various kinds of metadata which are mandatory, or could facilitate, the handling of different media types, including multimedia. The way in which the documents of different media types is used will be directly affected by the metadata. We classify metadata based on whether they are based on the data or information content of the documents. We also identify the type of metadata suitable for enabling correlation at a *semantic level*.

The extraction of metadata from data is highly influenced by the media type of the data. Our definition of **media type** is explained as follows. The American Heritage Dictionary defines *Medium* as a means of mass communication. This is generally determined by the way in which the information would be presented based on the original intention of the information creator. Of course, the same information could be stored in different physical formats and could be subjected to various transformations between creation and storage and between storage and presentation. We thus differentiate between what is meant by a data type, as used in conventional database technology, and media type, which depends on the presentation as opposed to the former. Issues of metadata extraction are discussed in Section 3.3. Metadata storage and organization is a very relevant issue and is discussed in Section 3.4.

### 3.1 Classification of metadata

In this section we review the different kinds of metadata used by researchers in different digital media types [KS94c]. We classify the various kinds of metadata based on whether they are based on the data or information content of the documents or not. The basic kinds of metadata we identify are:

- Content-dependent metadata.
- Content-descriptive metadata (Special case of Content-dependent metadata).
- Content-independent metadata.

Content-dependent metadata, as the name suggests, depends only on the content of the original data. When we associate metadata with the original data, which describes the contents in some way, but cannot be extracted automatically from the contents themselves, we call it content-descriptive metadata. This kind of metadata relates to characteristics, which could be determined exclusively by looking at the content (i.e., by the cognitive process), or derived intellectually with the support of tools and which could not have been derived on the basis of content alone.

A text index, like the document vectors in the LSI index [DDF+90] and the complete inverted WAIS index [KM91] are examples of content-based metadata. The index is determined by the content, e.g., the frequency and position of text units in the document.

Content-descriptive metadata, itself, can be classified as domain-dependent and domain-independent. Domain-dependent metadata uses domain-specific concepts. These concepts are used as a basis to determine the actual metadata created. An example of domain dependent metadata would characterize the set of images in a GIS database containing forest land cover. Domain-independent metadata, on the other hand, relies on no such domain-specific concepts. A typical example of a domain-independent metadata would be the one which describes the structure of a multimedia document [BR].

Content-independent metadata, on the other hand, does not depend on the content. This kind of metadata can be derived independently from the content of the data. This is like attaching a tag to the data irrespective of the data contents. Examples of content-independent metadata about a text-document are its date of creation and location.

Most of the work, so far, has concentrated on issues related to content-based and content-descriptive domain-independent metadata. These are not adequate for capturing the semantics of the domain. Content-descriptive domain-dependent metadata is needed to characterize the meaning and usage of the underlying objects.

We have proposed the use of content-descriptive domain-dependent metadata for structured data [KS94a]. We believe that the techniques for structured data can either be extended to or inspire analogous techniques for digital data. Image and visual data are inherently rich in semantic content. These objects can be better interpreted in the context of a given domain. Methods exist in advanced image analysis systems exist for low-level pattern recognition, image processing and segmentation and object recognition. Similarly various tools and techniques exist for other media types. We need to build upon these techniques to achieve correlation across different media types at a *semantic level* by associating these methods with content-descriptive metadata.

Jain and Hampapur [JH], have used video and image metadata for content-dependent access of videos in a video database. The Image and Video R-feature value pairs (e.g. the Object Track feature associated with a Set of image positions) may be considered as content-dependent metadata. The Image and Video Q-feature value pairs (e.g., the Video Class feature with associated values such as *News*, *Sports*) may be considered as content-descriptive metadata whereas the Meta feature value pairs (e.g., the Producer Info feature with the associated producer name) may be considered as examples of content-independent metadata.

Kiyoki et al. [KKH], have demonstrated a method for associating users' impressions of images with the images themselves. They create a semantic metadata space which is used to dynamically compute similarity between keywords and metadata items of the image. These may be considered as content-descriptive metadata.

Anderson and Stonebraker [AS] propose a metadata schema for satellite images. They lay emphasis on content-descriptive metadata for supporting temporal and geographic queries. These are primarily domain-independent.

Glavtisch et al. [GSW] discuss the use of subword units for speech documents for indexing. They use these indices to integrate speech documents in an information retrieval system. These may be considered as examples of content-dependent metadata.

Chen et al. [CHK$^+$] identify keyword locations, conversation segmentation by speaker and regions of speakers speaking emphatically, and index these for speech. Similarly they index features like keywords and layout for text image documents. These features are derived either in advance or at retrieval time. These may be considered as examples of content-dependent metadata.

Böhm and Rakow [BR] have suggested a classification of metadata for multimedia documents. The different types of metadata they identify and their relationship to our classification are:

- *Metadata for the Representation of Media Types.* This includes format, coding and compression techniques that may have been applied. These may be considered as content-independent metadata.
- *Content-descriptive Metadata.* A list of persons or institutions having some relation to a particular multimedia document's content are examples of these. We have also classified these metadata in a similar manner.
- *Metadata for Content Classification.* These refer to a classification of the content of a document. One way of classifying the content of a document is to identify the subject domain of the document. These are examples of domain-dependent content-descriptive metadata.
- *Metadata for Document Composition.* These refer to the logical components of multimedia documents. They may be considered as content-descriptive domain-independent metadata.

- *Metadata for Document History.* These metadata record the status of multimedia documents like approvedByEditor and notApproved. These may be considered as content-independent metadata.
- *Metadata for Document Location.* This may be considered as content-independent metadata.

A summary of the above discussion is presented in Table 2.

| **Metadata** | **Media Type** | **Content Dependence** |
|---|---|---|
| Q-Feature Value pairs (Jain and Hampapur) | Image, Video | content-descriptive |
| R-Feature Value pairs (Jain and Hampapur) | Image, Video | content-dependent |
| Meta Feature Value pairs (Jain and Hampapur) | Video | content-independent |
| Impression Vector (Kiyoki et al.) | Image | content-descriptive |
| Grid (Anderson and Stonebraker) | Image | content-dependent |
| Spatial Registration (Anderson and Stonebraker) | Image | content-descriptive |
| Temporal Information (Anderson and Stonebraker) | Image | content-independent |
| Speech feature index (Glavtisch et al.) | Audio | content-dependent |
| Keyword index (Chen et al.) | Text | content-dependent |
| Topic change indices (Chen et al.) | Audio | content-dependent |
| Layout indices (Chen et al.) | Image | content-dependent |
| Metadata for Representation of Media Types, Document History, Location (Böhm and Rakow) | MultiMedia | content-independent |
| Content Descriptive, Content Classification Document Composition Metadata (Böhm and Rakow) | MultiMedia | content-descriptive |

**Table 2.** Metadata Classification

Looking at the classification in Table 2, we observe that the Speech feature index and Impression vector are statistical correlations of the various terms of interest in the vocabulary. They do not represent semantic relationships. Also, R-feature value pairs, Grid and Metadata for Representation of Media Types, represent metadata influenced by the media type. Hence they cannot be used to correlate information *independent of the medium*. Spatial registration, keyword index, topic change indices, layout indices and metadata for document composition are domain-independent in nature. For semantic correlation, we should be able to capture domain-specific information independent of the medium of representation. This facilitates the representation of the meaning and use of the data in the documents. The metadata which satisfy this criteria are Q-feature value pairs and Content classification metadata. We believe that it is these types of metadata which provide the key for *semantic correlation*.

Our initial investigation on content-descriptive metadata which characterize the domain of GIS applications for the Site Planning Problem is discussed next.

### 3.2 Meta-correlation: The Key to Media-Independent Semantic Correlation

In this section we discuss with an example how related information represented in different digital media can be combined by making use of metadata. The relation between information in different media may be represented using *meta-correlations* as illustrated in Figure 4. In Section 3.1 we have identified the content-descriptive domain-dependent metadata as being suitable for information correlation across multiple representations. In this section we discuss an example from the GIS domain to illustrate information correlation across structured data and images. We can view the problem of correlating the information across different types of digital media from two perspectives:

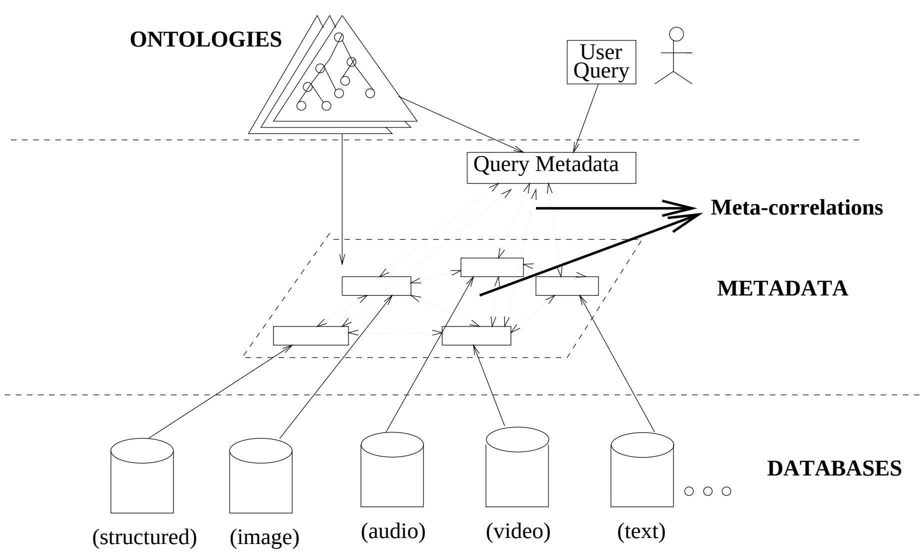


**Fig. 4.** Meta-correlations: Correlating Information using Metadata

**A Partial Schema for Digital Data** We can consider the metadata which captures the information represented in a particular media type as an *elementary schema* for the information. Consider an example of the GIS database which contains images on land use and land cover. One representation of the metadata for such a database is as follows.

```
[(region [ block (bounds [33N <= latitude <= 34N,
        84W <= longitude <= 85W])]) (relief [moderate, steep])
        (contiguous-area [large, medium])
        (shape [rectangular, square]) (land-use [urban, forest])]
```

This metadata entry for the image database states that all the images within it contain blocks with the following characteristics:

- All blocks fall within the latitudes 33N and 34N and longitudes 84W and 85W.
- All blocks have either moderate or steep relief.
- All blocks have large or medium contiguous areas.
- All blocks have rectangular or square shapes.
- All blocks are either of the urban or forest land use type.

This can be used as a starting point for representations of correlations with information represented in heterogeneous repositories. For instance, having identified areas of *moderate relief* one might establish a correlation with metadata associated with a structured database modeling population information for that region. This may then be considered as an example of a rudimentary *inter-schema correlation*.

The correlations described above can be exploited for browsing through related information represented in different media. For instance, in the example mentioned above, the user can decide to browse population information about a region after having determined its relief from the image database.

**Query Processing** The other perspective is to model the information need of a user as metadata guided by a domain-specific ontology. The *query metadata* then acts as the basis for correlation between the metadata for different digital media. Consider a query based on the browsing example discussed above.

Get me all regions having moderate relief and population greater than 200

The query metadata can be represented as follows:

```
[(region X) (population [> 200]) (relief [moderate])]
```

In this case the evaluation of the query results in computing the correlations in a dynamic manner at run time and can be processed as follows:

- The query metadata can be compared to the metadata corresponding to the structured database which has information population and retrieve the latitude and longitude values of all the areas having population greater than 200.
- The query metadata can be compared to the metadata corresponding to the image database. This results in invoking appropriate image processing routines and retrieving the images and the latitudes and longitudes of all areas having a moderate relief.
- The intersection of the latitude-longitude pairs can be computed. This correlation is illustrated in Figure 5.

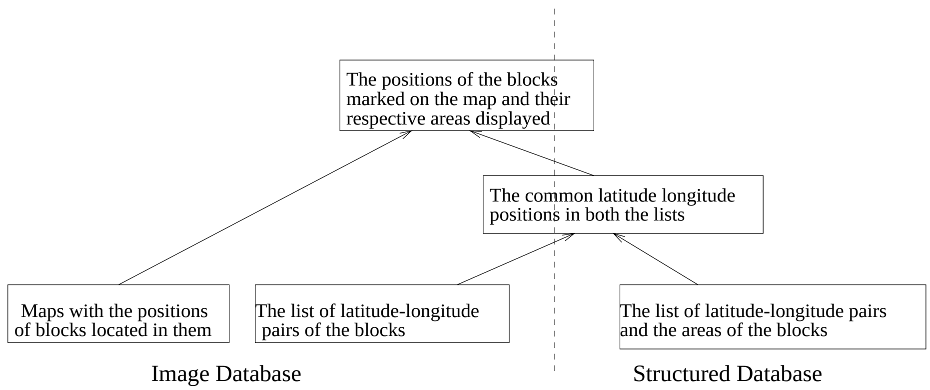


**Fig. 5.** Correlation of Structured and Image data using Metadata

### 3.3 Extractors for metadata

The information can be pre-processed to generate metadata or it could be subjected to access time metadata extraction. Extracting content-dependent metadata is entirely media type dependent. The extractors would automatically generate metadata based on the media type, e.g., an extractor for *TEXT* would filter out relevant words and index them. Metadata like sender, date, and subject could be generated from mail messages by an extractor for *MAIL* type which would look for lines starting with the keywords like *SUBJECT:* and *DATE:*, and return those lines. We could have extractors for *C* or *C++* [SSKS95] type which would recognize features such as functions, classes and sub-classes. The extractors for this type of metadata would require knowledge of the type of the underlying objects. They may also make use of magic tables that are reference tables which map patterns appearing in files or peculiar file names to data types.

Extracting content-dependent metadata for images would involve anticipating the range of user queries and is generally not feasible. Instead, some information like color and shapes could be extracted during pre-processing and others like patterns and outlines could be extracted at access time.

Content-independent metadata like size and location can be determined during pre-processing. We could view the media type itself as metadata. Metadata like container hierarchies for multimedia can either be explicitly supplied or extracted. Extraction of content-descriptive domain-dependent metadata involves associating semantics with the contents. The generation of such metadata involves automatic and semi-automatic approaches and is discussed at length in the previous section.

The extraction of any type of metadata depends on the range of user queries. Querying itself should then be independent of the metadata although the metadata could be used as a factor during querying. For example a query might utilize the size of the data as a retrieval criterion when transport costs have to be taken into account. Also, metadata could control the presentation and dynamic composition of retrieved information. If, based on content-descriptive metadata like type and size, the information is not presentable to the requester of the information, then, it would not be transported to the requester.

Jain and Hampapur [JH] have described various methods to extract content-dependent as well as content-descriptive metadata for their video database system. Content-dependent metadata, like the raw Image and Video Features, are extracted by the respective Feature Extractors which have low-level image and video processing routines. These would, for example, extract features like regions and lines from images. To generate content-descriptive domain-independent metadata, Image and Video Classifiers are employed which use a set of domain models to generate qualitatively labeled features for the image or video, like image brightness and texture. Users can also label, or annotate, images or videos with a unit called the Annotator. This would provide content-descriptive domain-dependent metadata. An Object Linker is used to maintain metadata regarding the temporal relationships between sets of frames, which denote a time-interval in video.

Kiyoki et al. [KKH] use different techniques to generate metadata for the orthogonal, semantic metadata space. The techniques relate to extracting content-descriptive metadata. The generation of domain-dependent metadata is done manually, where a small set of words are used to weigh an image. If a word corresponds to an image, as perceived by the metadata creator, a value of 1.0 is assigned for that word. Similarly, -1.0 is assigned if the word corresponds negatively and 0 is assigned otherwise. Domain-independent metadata is generated automatically by recognizing the colors in an image and using these to annotate the images based on some psychological models of correlating colors and words.

The extraction of metadata from satellite images [AS] follows the same basic principles. Content-dependent and content-descriptive domain-independent metadata are extracted automatically from the satellite images using operating system scripts and SQL triggers. Domain-dependent metadata, like keywords describing an image, are inserted manually into the database and associated with the respective images.

An interesting method of metadata generation is implemented by Grosky et al. [KS94c] to support intelligent browsing of structured media objects. User navigation patterns are captured and used to build relationships amongst the web of objects for individual users. This metadata is used for future user navigation. They call this metadata mediated browsing.

Glavitsch et al. [GSW] use speech recognition routines to extract sub-word features which are then used for indexing. Chen et al. [CHK+] employ a wide array of automatic metadata extractors. These locate user-specified keywords and phrases in images of text and audio streams. The located keywords and their locations are used as metadata. Other derived metadata include partition information in audio streams for different speakers, emphatic speech detection and sub-topic boundary locations.

### 3.4 Storage of metadata

Metadata can be stored in a variety of formats. The simplest would be as plain text files. Object-relational database systems could also be employed to manage the metadata. This would provide more flexibility with respect to querying the metadata itself and allowing the browsing of stored metadata prior to query building. Metadata in different media types could also be manipulated in an intuitive manner in such systems. We might relate unstructured data, e.g. images, and their structured data representations using metadata. In such cases it would be advantageous to store the metadata alongwith the original information. Metadata can be easily modified to reflect changes in the information contents or the content descriptions (e.g. location) if the metadata is stored in such database systems. Type specific functions could also be encapsulated alongwith the metadata to perform associative searches for unstructured information by using the metadata representing its features. Storing content-descriptive metadata, like container hierarchies, for multimedia objects determines the way in which the information can be browsed.

Query processing can be done off-line, as a pre-processing stage, if the vocabulary allowed in the query is of manageable size. Metadata based on this finite set of keywords and their locations could be pre-computed and stored for indexing. This is quite restrictive in most cases and searching provides better results if the vocabulary is unlimited. In this case metadata cannot be pre-computed and stored, but will have to be generated at run-time. Query processing on multimedia objects can be optimized if we store statistical metadata and metadata about logical structure.

Another issue in metadata storage is the location of the metadata for cases where the queries are submitted at a site remote to the site from where the information has to be retrieved. In most cases metadata is stored locally. Systems exist where metadata is stored at the remote sites, or could be pre-fetched for a query, and this is then used to intelligently analyze the query. For example, if we can determine from the metadata that the size of the query results could not be handled by the site where they are to be presented, appropriate action could be taken rather than retrieving the information and then failing.

## 4 Association of Digital Media Data with Metadata

As a result of query processing, the associated digital data also may be required to be displayed to the user (e.g., displaying the regions suitable for site location). Thus it is very important for associations to be stored between the extracted metadata and the underlying digital data. As discussed earlier, the type of metadata suitable for information correlation at a semantic level are the content-descriptive domain-dependent metadata. The main issues in associating these type of metadata to the actual data in the underlying digital media is to relate the domain-specific terms in the metadata to the domain-independent media-specific terms which might characterize the digital data.

### 4.1 Association of Metadata with Image Data

Consider once again our GIS application. Here we associate content-descriptive domain-dependent metadata with the underlying image data. This may be brought about by mapping the domain-dependent terms used to construct the metadata (e.g., moderate relief) to the domain-independent media-specific terms from the ontology (e.g., shape, color, texture, etc.). The mapping can be implemented as follows.

- Static embedding in an appropriate index/hash structure. For example, the value *green* for the meta-attribute[1] *vegetation* contains the qualitative spectral attribute *green* which could be mapped to a range of color coordinates (R,G,B) which conform reasonably to the notion of greenness (*wrt* the application domain). This can then be mapped to a set of images containing the regions of interest.
- The mapping of the meta-attribute to image object attributes could be done using a set of parameterized precompiled plans. The initial retrieval of the images containing the regions corresponding to the meta-attribute could be conducted by indexing/hashing on a subset of the image object attributes. The parametric plans can then be used to verify whether the mapping is successful by invoking the plan on the retrieved images and verifying whether the remaining image object attributes satisfy the constraints specified in the metadata.

### 4.2 Association of Symbolic descriptions with Image Data

We have seen the different kinds of content-dependent and content-descriptive metadata for images. We can perform semantic associative search on an image database if we could associate these two kinds of metadata. Kiyoki et al. [KKH] propose a "mathematical model of meaning" which provides functions for performing this kind of a search by using the metadata representing the image features. Here the abstract information about the images is used for their

[1] We are modeling domain-dependent content-descriptive metadata as a collection of meta-attribute value pairs.

indirect retrieval. In such systems, an orthogonal semantic space is maintained which could consist of the users' impressions as given by keywords with words describing their context and the image contents. The associative search is carried out in this orthogonal space.

### 4.3 Metadata for multimedia objects

When we consider multimedia as a separate media type, we need additional metadata beyond the ones we maintain for the component objects. We might have metadata associated with objects of media type text, images, audio or any other which we might see in the future, but a multimedia object will also have information about how these different objects are related to each other. This kind of metadata is not dependent on the contents of the constituent objects themselves, but we might think of it as being dependent on the content of the multimedia object itself. Thus, we could view these multimedia objects as composite objects and the relationship or structural metadata could be generated manually or automatically on the basis of pre-defined rules as in [SSB95].

We could also take a deeper view at this kind of metadata which holds information about the relationships between multimedia objects. When we think of multimedia objects as representing some real-world entities, we can say that the content of such objects are the real-world entities that they represent. This then becomes content-dependent metadata as we can use the media object to infer information regarding the real-world entity. Thus we might associate the identity of a person appearing in an image with the image. Grosky et al. [GFS] present a schema to support this notion. This schema allows intelligent associations between media objects and decides the way in which the user could browse through the database. Their architecture is capable of involving the user's navigation to form higher order clusters from this metadata using neural nets and genetic algorithms. This provides higher level concepts to the users which they can then modify. These clusters provide the user with a modified view of the metadata without actually modifying the metadata itself. Also, each user can maintain their own view of the metadata.

## 5 Conclusion

When the information relevant to a user resides in heterogeneous repositories and is stored in multiple representations, correlation of information at a higher *semantic level* may be required. As discussed in [Jai94], we also believe that correlation of information across various media types is possible only if we view them independent of their representation medium. We have presented a three level architecture comprising of the *ontology, metadata* and *data* to support such correlation. We used examples of a GIS application to explain components of this architecture.

We have also reviewed the use of vocabulary by various researchers to construct and design their metadata. We identify the domain-specific terms chosen with the application in mind as the most promising to support design and construction of metadata for semantic correlation. An important issue identified is the association of the metadata with the data stored in the various media. Typically this would involve relating the domain-specific, media-independent terms in the ontology to domain-independent, media-specific terms characterizing the data in a particular media type. We illustrate this with an example of the GIS domain mentioned above.

We have identified the *metadata level* as the most critical level in the three level architecture. The metadata should be able to model the *semantics* of the data which we characterize as the *meaning* and *use* of the data. We have reviewed the metadata designed for different media types by various researchers and analyzed them based on factors such as whether they model information specific to the domain of the data and whether they are specific to the media type. For the metadata to model the meaning of the data, it is important for it to capture as much domain-specific information as possible. Also, the metadata should be able to view the data independent of the medium of representation. Thus, we identify the domain-dependent, content-descriptive and media-independent metadata to be the best suited to support semantic correlation. We also discussed the type of metadata we support for semantic correlation in the example GIS application.

However for the state-of-the-art to overcome the "semantic bottleneck" [Jai94] the following research challenges should be met.

- The design of domain-dependent, media-independent metadata, and the constraints about the data that should be represented to capture the semantics of the data.
- The use of the metadata in comparing and combining information independent of the representation medium. This might involve combining and propagating constraints represented in the metadata for related data stored in different representation media.
- Determining a good set of terms and relationships among them to characterize the application domain and capture the semantic content of the data stored in the various media types.
- Determining a good set of terms that are media-specific and can characterize the information content of the data stored in that media type.
- Design of media-specific routines and indexing strategies to map the domain-dependent, media-independent terms to media-specific terms. This is important in the context of associating domain-specific metadata to the actual data.

**Acknowledgements** We thank Wolfgang Klaus for his collaboration in preparing the special issue [KS94c] on which some of the work in this chapter is based. The GIS example is based on ongoing collaboration with Dr. E. Lynn Usery at UGA.

## References


[AS] J. Anderson and M. Stonebraker. Sequoia 2000 Metadata Schema for Satellite Images, in [KS94c].

[BL+92] T. Berners-Lee et al. World-Wide Web: The Information Universe. *Electronic Networking: Research, Applications and Policy*, 1(2), 1992.

[BR] K. Böhm and T. Rakow. Metadata for Multimedia Documents, in [KS94c].

[CHK+] F. Chen, M. Hearst, J. Kupiec, J. Pederson, and L. Wilcox. Metadata for Mixed-Media Access, in [KS94c].

[CLT94] W.W. Chu, I.T. Leong, and R.K. Taira. A Semantic Modeling Approach for Image Retrieval by Content. *The VLDB Journal*, 3(4), October 1994.

[DDF+90] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Hashman. Indexing by Latent Semantic Indexing. *Journal of the American Society for Information Science*, 41(6), 1990.

[GFS] W. Grosky, F. Fotouhi, and I. Sethi. Content-based Hypermedia - Intelligent Browsing of Structured Media Objects, in [KS94c].

[Gru93] T. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition, An International Journal of Knowledge Acquisition for Knowledge-Based Systems*, 5(2), June 1993.

[GSW] U. Glavitsch, P. Schauble, and M. Wechsler. Metadata for Integrating Speech Documents in a Text Retrieval System, in [KS94c].

[GWJ91] A. Gupta, T. Weymouth, and R. Jain. Semantic Queries with Pictures: The VIMSYS Model. In *Proceedings of the 17th VLDB Conference*, September 1991.

[Jai94] R. Jain. Semantics in Multimedia Systems. *IEEE MultiMedia*, R. Jain, ed., 1(2), Summer 1994.

[JH] R. Jain and A. Hampapuram. Representations of Video Databases, in [KS94c].

[KKH] Y. Kiyoki, T. Kitagawa, and T. Hayama. A meta-database System for Semantic Image Search by a Mathematical Model of Meaning, in [KS94c].

[KM91] B. Kahle and A. Medlar. An Information System for Corporate Users : Wide Area Information Servers. *Connexions - The Interoperability Report*, 5(11), November 1991.

[KS94a] V. Kashyap and A. Sheth. Semantics-based Information Brokering. In *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM)*, November 1994. http://www.cs.uga.edu/LSDIS/infoquilt.

[KS94b] V. Kashyap and A. Sheth. Semantics-based Information Brokering: A step towards realizingthe Infocosm. Technical Report DCS-TR-307, Department of Computer Science, Rutgers University, March 1994. http://www.cs.uga.edu/LSDIS/pub.html.

[KS94c] W. Klaus and A. Sheth. Metadata for digital media. *SIGMOD Record, special issue on Metadata for Digital Media*, W. Klaus, A. Sheth, eds., 23(4), December 1994. http://www.cs.uga.edu/LSDIS/pub.html.

[KS95] V. Kashyap and A. Sheth. Semantic and Schematic Similarities between Databases Objects: A Context-based approach. Technical report, LSDIS Lab, University of Georgia (http://www.cs.uga.edu/LSDIS/infoquilt), January 1995.

[MS94] D. McLeod and A. Sheth. Interoperability in Multidatabase Systems. *Tutorial Notes - the 20th VLDB Conference*, September 1994.

[MS95] D. McLeod and A. Si. The Design and Experimental Evaluation of an Information Discovery Mechanism for Networks of Autonomous Database Systems. In *Proceedings of the 11th IEEE Conference on Data Engineering*, February 1995.

[She91] A. Sheth. Semantic issues in Multidatabase Systems. *SIGMOD Record, special issue on Semantic Issues in Multidatabases*, A. Sheth, ed., 20(4), December 1991. http://www.cs.uga.edu/LSDIS/pub.html.

[SK92] A. Sheth and V. Kashyap. So Far (Schematically), yet So Near (Semantically). *Invited paper in Proceedings of the IFIP TC2/WG2.6 Conference on Semantics of Interoperable Database Systems, DS-5*, November 1992. http://www.cs.uga.edu/LSDIS/pub.html.

[SSB95] L. Shklar, K. Shah, and C. Basu. The InfoHarness Repository Definition Language. In *Proceedings of the Third International WWW Conference*, May 1995.

[SSKS95] L. Shklar, A. Sheth, V. Kashyap, and K. Shah. Infoharness: Use of Automatically Generated Metadata for Search and Retrieval of Heterogeneous Information. In *Proceedings of CAiSE '95*, June 1995. http://www.cs.uga.edu/LSDIS/infoharness.

[TC94] P. Tsai and A. Chen. Concept Hierarchies for Database Integration in a Multidatabase System. In *Advances in Data Management*, December 1994.

[Wie94] G. Wiederhold. Interoperation, Mediation and Ontologies. In *FGCS Workshop on Heterogeneous Cooperative Knowledge-Bases*, December 1994.

[YSDK91] C. Yu, W. Sun, S. Dao, and D. Keirsey. Determining relationships among attributes for Interoperability of Multidatabase Systems. In *Proceedings of the 1st International Workshop on Interoperability in Multidatabase Systems*, April 1991.

This article was processed using the LaTeX macro package with LLNCS style